2019-09-09

The redundancy effect is related to a lack of conditioned inhibition: Evidence from a task in which excitation and inhibition are symmetrical

Zaksaite, Gintare

10.1177/1747021819878430
Quarterly Journal of Experimental Psychology
SAGE Publications

The redundancy effect is related to a lack of conditioned inhibition: Evidence from a task in which excitation and inhibition are symmetrical

Tara Zaksaite and Peter M. Jones

School of Psychology, University of Plymouth, Plymouth, UK

Author note

The authors declare that there is no conflict of interest.

Correspondence concerning this article should be addressed to: Tara Zaksaite, School of Psychology, Portland Square Building, University of Plymouth, Plymouth, PL4 8AA, UK. Email address: tara.zaksaite@plymouth.ac.uk.

Funding

This work was conducted as part of Tara Zaksaite's Ph.D., funded by the University of Plymouth.

Abstract

Rescorla and Wagner's (1972) model of learning describes excitation and inhibition as symmetrical opposites. However, tasks used in human causal learning experiments, such as the allergist task, generally involve learning about cues leading to the presence or absence of the outcome, which may not reflect this assumption. This is important when considering learning effects which provide a challenge to this model, such as the redundancy effect. The redundancy effect describes higher causal ratings for the blocked cue X than for the uncorrelated cue Y in the design A+/AX+/BY+/CY-, the opposite pattern to that predicted by the Rescorla-Wagner model, which predicts higher associative strength for Y than for X. Crucially, this prediction depends on cue C gaining some inhibitory associative strength. In this manuscript, we used a task in which cues could have independent inhibitory effects on the outcome, to investigate whether a lack of inhibition was related to the redundancy effect. In Experiment 1, inhibition for C was not detected in the allergist task, supporting this possibility. Three further experiments using the alternative task showed that a lack of inhibition was related to the redundancy effect: the redundancy effect was smaller when C was rated as inhibitory. Individual variation in the strength of inhibition for C also determined the size of the redundancy effect. Given that weak inhibition was detected in the alternative scenario but not in the allergist task, we recommend carefully choosing the type of task used to investigate associative learning phenomena, as it may influence results.

Key words: inhibition, allergist task, redundancy effect, human causal learning, Rescorla-Wagner (1972) model

The redundancy effect is related to a lack of conditioned inhibition: Evidence from a task in which excitation and inhibition are symmetrical

In the study of associative learning, it is generally acknowledged that animals are capable of learning at least two kinds of relationships between events. The first kind of relationship is learned when the occurrence of a cue indicates that an outcome is *more* likely (excitation), and the second is learned when the cue indicates that the outcome is *less* likely (inhibition). Pavlov (1927) is perhaps best known for his demonstrations of the first kind of learning, but he was also the first to demonstrate learning of an inhibitory relationship (see also Konorski, 1948). In a conditioned inhibition procedure, animals first learn that a single cue causes the outcome (A+). On subsequent trials, the addition of a second cue results in the omission of the outcome (AB^0). B becomes an inhibitor of the outcome because it signals the absence of the outcome that occurred when A was presented alone. The inhibitory properties of B can be observed in several ways (Rescorla, 1969), most notably a summation test. This involves the presentation of the inhibitor, B, with a further cue that has been separately paired with the outcome (e.g. C+). If B has become an inhibitor then a weaker conditioned response to the compound BC should be observed than for C alone. Likewise, in studies of human causal learning, participants should report a lower expectation of the outcome on BC trials (e.g. Dickinson, Shanks, & Evenden, 1984).

Despite widespread demonstrations of both conditioned excitation and inhibition, there is little consensus as to the conceptual relation between the two. The popular theory proposed by Rescorla and Wagner (1972) describes these two kinds of learning simply as opposites of one another. According to their theory, the unexpected occurrence of the outcome results in a positive prediction error and excitatory learning. The unexpected omission of the outcome, however, results in a negative prediction error and inhibitory learning. In the example above, a positive association between A and the outcome should be

formed, leading to a negative prediction error on AB^0 trials, and consequently an inhibitory association between B and the outcome. Rescorla and Wagner's model is therefore symmetrical in its conception of excitation and inhibition. One consequence of this symmetry is that it predicts equivalent effects of presenting a cue in the absence of any outcome, whether the cue is an excitor or an inhibitor. That is, it predicts that extinction of these two kinds of learning should be similar. This prediction was first tested by Zimmer-Hart and Rescorla (1974), who were unable to find any evidence that inhibition could be extinguished by presenting the inhibitory cue alone. This differs from the finding that excitation can be readily extinguished by nonreinforced presentation, and led Zimmer-Hart and Rescorla to conclude that inhibition might not be the symmetrical opposite of excitation. Subsequent tests of extinction of inhibition have found mixed results, both in humans (Yarlas, Cheng, & Holyoak, 1995) and non-human animals (e.g. Detke, 1991; Holland, 1985; Miller & Schachtman, 1985; Pearce, Nicholas, & Dickinson, 1982; Rescorla, 1982; Williams & Overmier, 1988). The failure to find reliable evidence that inhibition operates similarly to excitation is cited as a notable failure of Rescorla and Wagner's model (Miller, Barnet, & Grahame, 1995).

An alternative view, proposed by Zimmer-Hart and Rescorla (1974), is that inhibitory cues do not acquire associative strength that is the simple opposite of the association held by the excitor. Instead, they suggested that inhibitors act by raising the threshold at which the outcome is expected; inhibitors are not expected to have any independent effects when presented alone, and such presentations should not cause extinction of inhibition to occur. Zimmer-Hart and Rescorla's theory therefore, incorporates a different assumption about the nature of inhibition to Rescorla and Wagner's (1972) theory. The evidence discussed so far seems to be most consistent with Zimmer-Hart and Rescorla's theory, but there is some evidence from studies of human causal learning that this depends on the exact task given to

participants. In particular, whether or not extinction of inhibition is observed appears to depend on whether the task incorporates the assumptions of Rescorla and Wagner's theory, or Zimmer-Hart and Rescorla's. Melchers, Wolff, and Lachnit (2006; see also Lotz & Lachnit, 2009) found extinction of an inhibitory cue when reinforcers could take on negative values rather than signalling an absence of the outcome that would have otherwise occurred. In Melchers et al. study, participants were trained with one of two scenarios. In both, participants were asked to learn which foods influenced the level of a hormone in a hypothetical patient. For one group, the outcome was binary and the hormone level either increased or stayed the same on each trial. For the other group, there were three possible levels of the outcome: an increase in the hormone level, no change, or a decrease. Extinction of inhibition was observed in this three-outcome group, but not in the group trained with the binary outcome. The authors hypothesised that the reason for these results was that the scenario with three outcome levels generated an expectation that decreases in the outcome would occur when an inhibitory cue was presented alone. When an inhibitory cue was presented alone and led to no effect, extinction of inhibition occurred. In the two-outcome scenario, however, reinforcers varied only unidirectionally. Therefore, decreases in the level of the outcome should not have been expected when the inhibitor was presented alone, and no extinction of inhibition occurred. A slightly different interpretation of the results was proposed by Baetu and Baker (2010), who suggested that the critical difference between the two groups was not that the outcomes could take on negative values, but that the groups received different instructions and were asked to rate the cues using different rating scales. For the two-outcome group, participants were told that negative values on the rating scale prevented an increase in hormone levels, in accordance with the threshold conception of inhibition favoured by Zimmer-Hart and Rescorla. For the three-outcome group, negative values on the rating scale indicated that the cue decreased hormone levels, more consistently

with the manner of inhibition described by Rescorla and Wagner, where inhibitors would have the opposite effects to excitors. Regardless of which interpretation is correct, it seems that the predictions of Rescorla and Wagner's model are a better match for the data when the task used matches the assumptions of the model.

Following this line of reasoning, other causal learning effects related to inhibition which are discrepant with the predictions of Rescorla-Wagner (1972) model might also be constrained by the properties of the task. One such recently-observed result is the redundancy effect (Uengoer, Lotz, & Pearce, 2013; for analogous results in non-human animals, see Pearce, Dopson, Haselgrove, & Esber, 2012). Uengoer et al. trained participants to predict whether or not a fictional patient would suffer a stomach ache when he consumed various different foods. Training consisted of four trial types: A+, AX+, BY+, and CY^0, where each letter represents a food and "+" and "^{0}" represent the presence and absence of the stomach ache, respectively. Following this training, participants were asked to rate the probability of the stomach ache for each food alone. Participants indicated that the stomach ache was more likely to occur if the patient consumed X than if he consumed Y. This finding is referred to as the 'redundancy effect' (Jones & Pearce, 2015), and it poses a challenge to models of learning which compute prediction error in the way described by Rescorla and Wagner (1972). Their model predicts that Y, although irrelevant to the solution of the BY+/CY^0 discrimination, should nonetheless become a moderate excitor for stomach ache. This is because it gains excitatory associative strength on BY+ trials, and is protected from extinction on CY^0 trials to some extent by the acquisition of inhibitory strength by C. By contrast, X should be 'blocked' by A on AX+ trials (as in Aitken, Larkin, & Dickinson, 2000; Kamin, 1969), and should have little associative strength by the end of training. The model therefore predicts that Y should be a stronger excitor than X, which is the opposite result to that which Uengoer et al. observed. Crucially, this prediction depends on C becoming an

inhibitor for stomach ache. If C does not become an inhibitor, Y is not protected from extinction and should not retain an excitatory association with the outcome. As we have already described, however, the accuracy of Rescorla and Wagner's model regarding inhibition is dependent on whether or not the task reflects the assumptions of the model. It is notable that in Uengoer et al.'s experiments, and in all other published demonstrations of the redundancy effect in humans (Jones & Zaksaite, 2018; Jones, Zaksaite, & Mitchell, 2019; Uengoer, Dwyer, Koenig, & Pearce, 2019; Zaksaite & Jones, 2017), only positive and neutral outcomes were presented. The instructions given to participants in Uengoer et al.'s (2013) experiments asked them to learn whether the consumption of different foods led to stomach ache or not (p. 325), which may have indicated that these were the only two relationships with the allergic reaction that were possible. This is in contrast to prior demonstrations of conditioned inhibition using an allergist task, in which the possibility of foods preventing an allergic reaction was made explicit (e.g. Larkin, Aitken, & Dickinson, 1998; Melchers, Lachnit, & Shanks, 2004). We are aware of only one demonstration of conditioned inhibition using a food-allergy task in the absence of instructions that some foods will prevent the allergic reaction (Karazinov & Boakes, 2004), and this used a migraine outcome rather than the stomach ache used by Uengoer et al. Furthermore, conditioned inhibition might be particularly difficult to obtain using an allergist task because real-world experience of how foods cause aversive reactions is likely to contain few examples of foods preventing allergic reactions that would otherwise occur. It seems reasonable, therefore, to consider whether the redundancy effect might be due at least in part to participants' failure to learn that C prevents stomach ache.

In Experiment 1, we tested this idea by training participants in a similar way to Uengoer et al. (2013), and subsequently assessing whether or not C had become an inhibitor for stomach ache in a summation test. To foreshadow our results, we found no evidence that

participants learned that C was inhibitory. In subsequent experiments we used a task in which outcome levels could decrease and independent inhibitory effects on the outcome could be observed. In such a task we may expect to observe greater inhibition for C, and consequently greater excitation for Y than for X. Such findings would bring us closer to reconciling Uengoer et al.'s findings with Rescorla and Wagner's (1972) model.

Experiment 1

In Experiment 1 we explored whether evidence of inhibition for C would be obtained in an allergist task, identical to the ones used in the previous studies on the redundancy effect in humans (Uengoer et al., 2013; Jones et al., 2019; Jones & Zaksaite, 2018; Zaksaite & Jones, 2017). In a design which included the trial types A+/AX+/BY+/CY^0, participants were asked to learn which foods caused a stomach ache in a fictional patient. The outcome in this task was a stomach ache which could either occur or not occur. If we demonstrated the redundancy effect but no inhibition for C, this would indicate that a lack of inhibition could be related to the redundancy effect. If we demonstrated the redundancy effect alongside inhibition for C, this would indicate that the redundancy effect is likely due to other factors than a lack of inhibition for C. The full design of Experiment 1 is presented in Table 1.

To enable us to check whether C gained inhibitory associative strength, the design of Experiment 1 included a neutral cue G (GH^0). In addition, because the Rescorla-Wagner (1972) model predicts that inhibition for C will be weak, we also included cue E, trained as an inhibitor ($D+/DE^0$). This enabled us to compare whether C gained inhibitory associative strength, and whether its inhibition was as strong as for a cue trained as an inhibitor. At test, participants were presented with three compounds involving cues C, the inhibitory cue E, and the neutral cue G. Each of these cues was presented with cue F, which was trained as an excitor (F+), forming compounds CF, EF, and GF. This summation test aimed to measure the

extent to which cues C, E, and G reduced responding to the excitatory cue, as per recommendations of Rescorla (1969).

Table 1.

The design of Experiment 1. Letters represent different cues, "+" denotes trials leading to a stomach ache, and "^{0}" denotes trials leading to no stomach ache.

Stage 1	Test
A+	A
AX+	B
BY+	C
CY^0	D
D+	E
DE^0	F
F+	G
GH^0	H
	X
	Y
	CF
	EF
	GF
x 16	x 2

Method

Participants. Participants were 33 University of Plymouth students studying Psychology. They received course credit for participation in this experiment. They were aged 18-26 years (M = 19.67, SD = 1.8) and four were male. They were tested in individual

cubicles. The sample size was chosen to be similar to the previous behavioural research related to the redundancy effect (e.g. Uengoer et al., 2013).

Materials. The experiment was presented on a 22-inch desktop computer with a 1920 x 1080 screen resolution. The experiment was designed, presented, and responses were recorded using E-prime 2.0 software (Psychology Software Tools, PA, US).

The cues were 10 images of foods on a white background, 300 x 300 pixels. The foods were: apple, banana, broccoli, cabbage, cherries, grapes, orange, pumpkin, strawberries, and watermelon. The foods were randomly assigned to each type of cue (A, B, C, D, E, F, G, H, X, Y) for each participant. The outcomes were stomach ache, signified by text and an image of a sad face on a red background, and no stomach ache, indicated by text and an image of a happy face on a green background. The stimuli and outcomes were presented on the screen with a black background and white text.

Procedure. The instructions and procedure for the learning task were adapted from Uengoer et al. (2013) and the parameters of the task were consistent with other experiments on the redundancy effect in our laboratory. Initial instructions were presented on the screen as follows:

This study is concerned with the question of how people learn about relationships between different events. In the present case, you should learn whether the consumption of certain foods leads to stomach ache or not.

Imagine that you are a medical doctor. One of your patients often suffers from stomach ache after meals. To discover the foods the patient reacts to, your patient eats specific foods and observes whether stomach ache occurs or not.

The results of these tests are shown to you on the screen one after the other. You will always be told what your patient has eaten. Sometimes he has only consumed a single kind of food, and other times he has consumed two different foods. Please look at the foods carefully.

Thereafter you will be asked to predict whether the patient suffers from stomach ache. For this prediction, please click on the appropriate response button. After you have made your prediction, you will be informed whether your patient actually suffered from stomach ache.

Use this feedback to find out what causes the stomach ache your patient is suffering from. Obviously at first you will have to guess because you do not know anything about your patient, but eventually you will learn which foods lead to stomach ache in this patient and you will be able to make correct predictions.

For all of your answers, accuracy rather than speed is essential. Please do not take any notes during the experiment.

If you have any questions, please ask them now. If you do not have any questions, please start the experiment by clicking the mouse.

In Stage 1 participants were presented with 16 blocks of trials. The eight trial types (A+, AX+, BY+, CY^0, D+, DE^0, F+, GH^0) appeared once per block in a random order. There were no successive repetitions of the same trial type. Each trial started with the presentation of either one or two images of foods in the top half of the screen, below the phrase "The patient ate the following food:" (or "The patient ate the following foods:" for trials with two images). For trials with two images, one was located on the left and one on the right (counterbalanced), while images of single foods were located in the middle. The sentence "Which reaction do you expect?" was presented below the images. Participants responded by clicking one of two response buttons placed at the bottom of the screen. The left-hand button was labelled "No stomach ache" and the right-hand button was labelled "Stomach ache". As soon as the participant responded, the response buttons and the sentence above them were replaced by a statement and picture showing the outcome of the trial, while the images of the cues and the sentence "The patient ate the following foods:", remained. When the outcome was stomach ache, the statement was "The patient has stomach ache" and the picture of the sad face on a red background was shown. When the outcome was no stomach ache, the statement was "The patient has no stomach ache", and the picture of the happy face on a green background was shown. This feedback display remained on the screen for 3000 ms, followed by a 500 ms blank screen after which the next trial began.

After all of the trials in Stage 1 were completed, participants were shown the following instructions:

Now, your task is to judge the probability with which specific foods cause stomach ache in your patient. For this purpose, foods will be shown to you on the screen.

In this part, you will receive no feedback about the actual reaction of the patient. Use all the information that you have collected up to this time.

The test stage then began. On each trial, the sentence "What is the probability that the food causes stomach ache?" was shown above an image of a single food or two images of foods. Participants responded by clicking on an 11-point rating scale ranging from 0 (*Certainly not*) to 10 (*Very certain*). The rating scale was located in the lower half of the screen, oriented horizontally. After each response, a blank screen was shown for 500 ms and was followed by the next trial. At test participants were asked to rate the single foods as well as three compounds of two cues (CF, EF, GF) twice; the trial types were presented in a random order. For each participant, the average of the two causal ratings for each type of trial was calculated and used in the analyses.

Data Analyses. To investigate whether there were any significant differences between ratings for the cues, Analysis of Variance (ANOVA) tests and t-tests were performed on the data. When paired comparisons between more than two levels of a factor were made, Bonferroni corrections were used. The alpha level of significance was set at .05 for all other comparisons including when using t-tests for the comparisons of interest (e.g. the redundancy effect, inhibition for C) and simple main effects analyses. When data violated the assumption of sphericity, Greenhouse-Geisser corrections were applied to degrees of freedom. For null comparisons of interest, Bayes Factors (BF_{01}) were calculated using a JZS prior with a scaling factor of 0.707. To calculate Bayes Factors, JASP version 0.6 was used (JASP Team, 2015). Bayes Factors greater than three are considered to provide support for the null hypothesis. Bayes Factors less than one-third indicate support for the alternative hypothesis (Jeffreys, 1961). Estimates of effect sizes reported for ANOVA tests were eta squared (η^2) and partial eta squared (η_P^2). Effect-size estimates for t-tests were Cohen's d_s for between-

subjects comparisons and Cohen's d_z for within-subjects comparisons, in accordance with recommendations by Lakens (2013).

Ethical approval. The ethical approval for the experiments detailed in this manuscript was granted by the University of Plymouth, Faculty of Health and Human Sciences Research Ethics Committee and all participants provided informed consent in writing before taking part in the study.

Results

Stage 1. Figure 1 shows proportions of stomach-ache predictions throughout the eight epochs of Stage 1. For each participant, an epoch was defined as an average of responses on two successive trials of the same trial type. This figure indicates that participants learned the contingencies. In the final epoch they responded correctly on 99.22% (*SD* = 6.21%) of the trials.

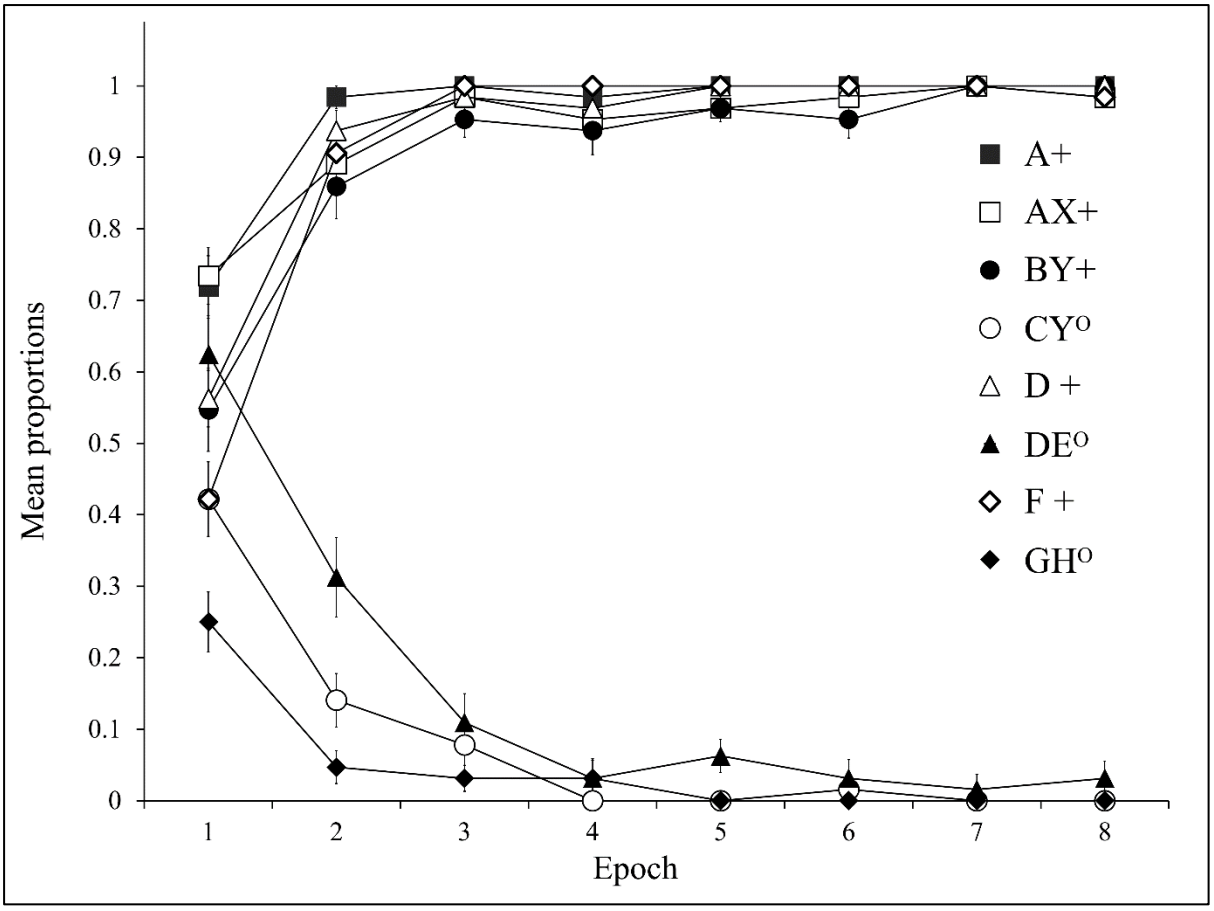

Figure 1. Mean proportions of stomach-ache predictions throughout the eight epochs of Stage 1 in Experiment 1. The error bars on this and the subsequent figures show the standard error of the mean, adjusted to exclude between-subjects variability as recommended by Cousineau (2005), unless stated otherwise.

Test. Figure 2 shows causal ratings for each trial type at test. Ratings for G and H were higher ratings than all other cues, ts ≥ 3.96, ps $\leq .014$, d_zs $\geq .69$. Cues C, E, and G/H had lower ratings than all other cues, ts ≥ 3.66, ps $\leq .032$, d_zs $\geq .64$. Cues X and Y had intermediate ratings, which were different from those of all other cues, ts ≥ 3.66, ps $\leq .032$, d_zs $\geq .64$. A t-test indicated that the redundancy effect was observed: X had significantly higher ratings than Y, $t(32) = 3.94$, $p < .001$, $d_z = .69$.

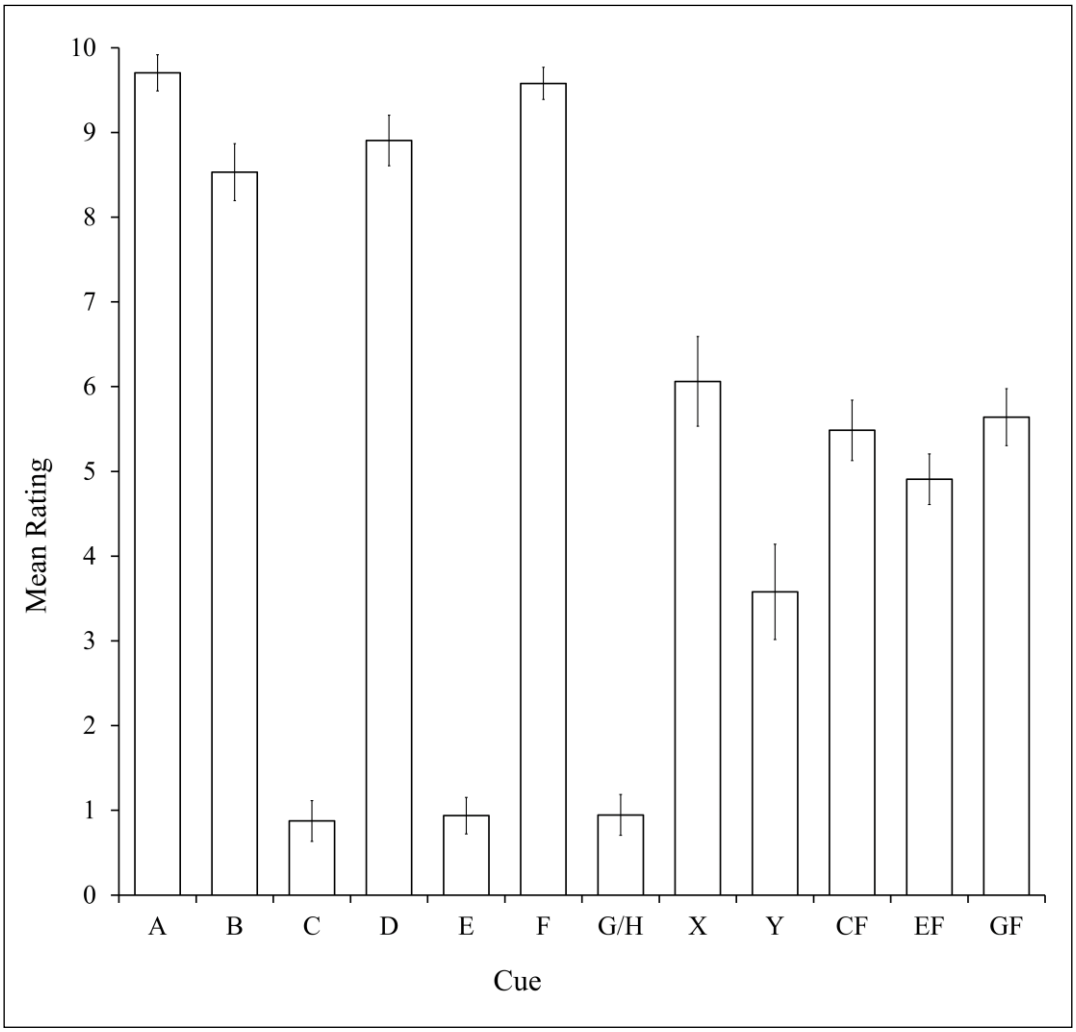

Figure 2. Mean causal ratings at test in Experiment 1 (±SEM).

Regarding the summation tests, a one-way ANOVA for trial type (CF, EF, HF) indicated that ratings between CF, EF, and GF did not differ significantly, $F(2, 64) = 1.13$, $p = .33$, $\eta^2 = .03$, $BF_{01} = 4.46$. Paired t-tests indicated that C did not become inhibitory as ratings for CF did not differ significantly from ratings involving a neutral cue, GF, $t(32) = .49$, $p = .63$, $d_z = .09$, $BF_{01} = 4.81$. One might expect that the best evidence for inhibition, if present, would be demonstrated by the comparison between EF and GF. A paired t-test indicated no significant differences between ratings for EF and GF, $t(32) = 1.4$, $p = .172$, $d_z = .24$, $BF_{01} = 2.22$, suggesting a failure to obtain inhibition in this experiment. However, the

Bayes Factor was less than 3 for this comparison and thus did not provide evidence for the null result.

Discussion

In Experiment 1, we explored whether evidence of inhibition for C, and the redundancy effect could be obtained in an allergist task. While the redundancy effect was observed, we did not find evidence of inhibition for C, despite showing that participants learned the contingencies. Therefore, it is possible that a lack of inhibition for C contributed to the redundancy effect in this experiment. In addition to this, we did not find inhibition for E, which was trained as an inhibitor. This has implications for research investigating cues which are predicted to become inhibitory in an allergist task. It suggests that obtaining inhibition may be difficult, particularly for cues which are predicted to be weak inhibitors, such as C. One possible reason for this could have been the asymmetry between inhibition and excitation in this task, consistently with suggestions by Melchers et al. (2006). We examined this possibility in the next experiments by using an alternative task in which outcome levels could decrease as well as increase. This task aimed to encourage participants to interpret inhibition consistently with the assumptions of the Rescorla-Wagner (1972) model. The task is described in detail in the next section.

Experiment 2

The task used in Experiment 2 and the subsequent experiments, involved asking participants to learn about a fictional patient who consumed medicines which could lead to an increase, no change, or a decrease in the levels of a fictional hormone. An increase in hormone levels represented excitatory effects on the outcome, no change in hormone levels represented neutral effects on the outcome, and a decrease in hormone levels represented inhibitory effects on the outcome.

Firstly, we aimed to see whether evidence of inhibition would be obtained in this task, and in particular, inhibition for C. Secondly, we aimed to see whether the redundancy effect would be observed, or whether it would be reversed, in line with predictions of the Rescorla-Wagner (1972) model. The design of Experiment 2 is presented in Table 2.

Table 2.

The design of Experiment 2. Letters represent different medicines, "+" refers to an increase, "^{0}" to no change, and "-" to a decrease in hormone levels.

Stage 1	Test
A+	A
AX+	B
BY+	C
CY^0	D
D+	E
DE^0	F
F+	G
G-	H
H^0	X
	Y
	CF
	EF
	HF
x 20	x 2

To check whether inhibition was obtained, E was established as an inhibitory cue, leading to a decrease in hormone levels ($D+/DE^0$). Cue H was shown to be neutral and led to no change in hormone levels (H^0). Participants were also presented with a single cue which

led to a decrease in hormone levels (G-) to make sure they saw evidence that single cues could have independent inhibitory effects on the outcome. In order to test whether inhibition was obtained, once again a summation test was used with a causal transfer cue F (F+). At test, F was paired with C, the inhibitory cue E, and the neutral cue H, forming three compounds: CF, EF, and HF. To test whether inhibition occurred, ratings for EF and for HF were compared. To test whether C became an inhibitor, ratings for CF and for HF were compared. To test whether C became as strong an inhibitor as E, ratings for CF and EF were compared.

Method

Participants. Participants were 31 University of Plymouth students aged 18-62 years (M = 27.06, SD = 12.63); 12 were male.

Materials. The materials and procedure in Experiment 2 were the same as in Experiment 1 unless otherwise stated.

The stimuli were 10 images of different colour medicines on a white background, 300 x 300 pixels. Images of the medicines were: brown, green, magenta, orange, pink, purple, red, turquoise, white, and yellow. The medicines were randomly assigned to each type of cue (A, B, C, D, E, F, G, H, X, Y) for each participant. The levels of the outcome were: an increase, signified by text “The level of hormone increased” and an image of a yellow arrow pointing upwards on a white background; no change, indicated by text “The level of hormone did not change” and an image of a grey-horizontal arrow pointing left and right on a white background; a decrease, indicated by text “The level of hormone decreased” and an image of a blue arrow pointing downwards on a white background.

Procedure. The instructions for the learning task were adapted from Experiment 1, and were presented on the screen as follows:

Imagine that you are a medical researcher, interested in the effects of different medicines on hormone levels. Your task is to figure out whether the consumption of different medicines will result in an increase, no change, or a decrease in hormone levels. Sometimes one medicine will be consumed and sometimes two medicines will be consumed together.

In the cases where consuming two medicines leads to no change in hormone levels, one medicine may cause an increase and the other a decrease in hormone level, cancelling each other's effects out. However it is also possible that both of the medicines lead to no change.

On the following screens you will see the medicines that participants consume and will be asked to predict whether hormone level will increase, decrease, or will not change by clicking the corresponding button. Then, you will be informed of the resulting hormone level change, if any.

At the beginning you will have to guess but by using the feedback provided your guesses should become more accurate. Accuracy is more important than speed for your answers; you may take as long as you like on each trial.

If you have any questions, please ask the experimenter now. Alternatively please click the mouse to start the experiment.

In Stage 1 participants were presented with 20 blocks of trials, with each of the nine trial types (A+, AX+, BY+, CY^0, D+, DE^0, F+, G-, H^0) occurring once per block. As in the previous experiment, the order of the trial types was random with no successive repetitions of the same trial type. On each trial the screen displayed either one or two images of medicines with the caption "The following medicine was administered" (or "The following medicines were administered" for trials with two medicines). The sentence "What effect on hormone level do you expect?" was presented below the images. Participants responded by clicking one of three response buttons placed at the bottom of the screen. The left-hand button was labelled "Decrease", the middle button "No change", and the right-hand button was labelled "Increase". As soon as participants responded, the screen was replaced by a statement and an image showing the outcome of the trial. When the outcome was an increase, the statement was "The level of hormone increased" and the image was a yellow arrow pointing upwards. When the outcome was no change, the statement was "The level of hormone did not change" and the image was a grey-horizontal arrow. When the outcome was a decrease, the statement was "The level of hormone decreased" and the image was a blue arrow pointing downwards.

Similarly to the previous experiment, predictions on each epoch were calculated as an average of two responses for each trial type, and averaged across participants.

Test instructions were as follows:

Next, you are asked to provide your final report. Medicines will be presented on the screen and your task is to use all of the information you have collected up to this time to judge the probability to which specific medicines will change hormone level. Please rate them on a scale from "Decrease" to "Increase" by clicking on the corresponding button.

You will receive no feedback about the resulting hormone level in this stage.

Please click the mouse to begin.

On each test trial either one or two medicines were presented on the screen. Above the image(s) was the sentence "Following the consumption of this medicine the level of hormone will:" (or "Following the consumption of these medicines the level of hormone will:" for trials with two images). Participants responded by clicking on a 21-point horizontal rating scale ranging from -10 (*Decrease*) through 0 (*Not change*) to 10 (*Increase*). Similarly to the previous experiment, each trial type at test was presented twice with no successive repetitions, and the average of the two ratings was used in the analyses.

Results

Stage 1. Predictions of hormone-level changes in Stage 1 are displayed in Figure 3. This figure shows that participants were able to learn the pairings. On the final epoch they responded correctly on 96.42% (*SD* = 16.28%) of the trials.

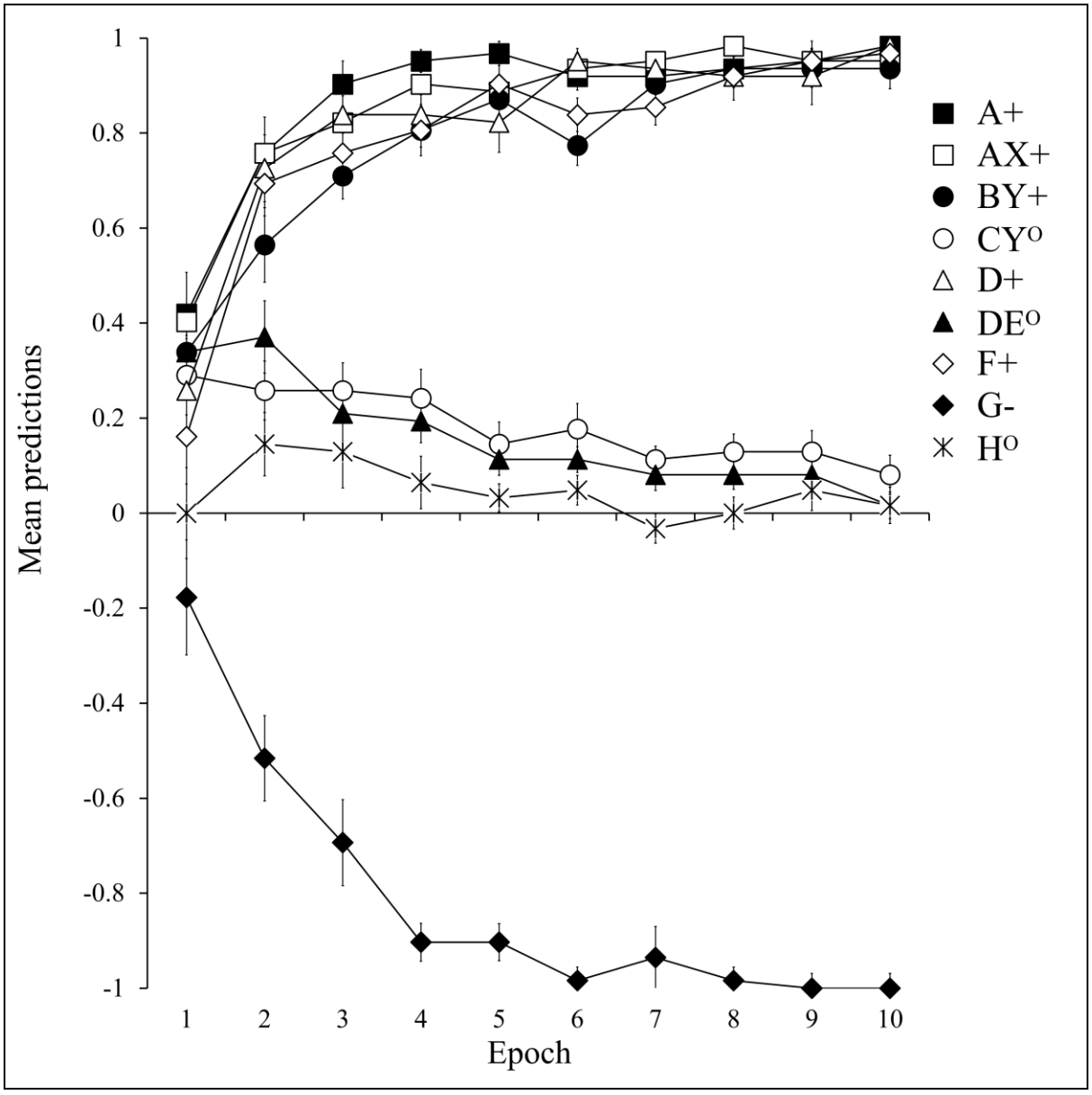

Figure 3. Mean hormone-change predictions throughout the ten epochs of Stage 1 in Experiment 2 (±SEM). Positive values refer to a prediction of an increase, zero values refer to a prediction of no change, and negative values refer to a prediction of a decrease in hormone levels.

Test. Figure 4 shows mean causal ratings at test. A one-way ANOVA comparing the cues (A, B, C, D, E, F, G, H, X, Y) revealed a significant main effect, $F(3.54, 106.05) = 97.43$, $p < .001$, $\eta^2 = .77$. Bonferroni-corrected paired comparisons indicated that ratings for A, D, and F were higher than for all other cues, ts ≥ 3.52, ps $\leq .062$, d_zs $\geq .63$. Ratings for E

were lower than all other cues except for C, $ts \geq 5.4$, $ps < .001$, $d_z s \geq .97$, and ratings for G were lower than for all other cues, $ts \geq 5.96$, $ps < .001$, $d_z s \geq 1.07$. A paired t-test indicated that the redundancy effect was obtained; ratings for X were higher than for Y, $t(30) = 2.42$, $p = .022$, $d_z = .43$.

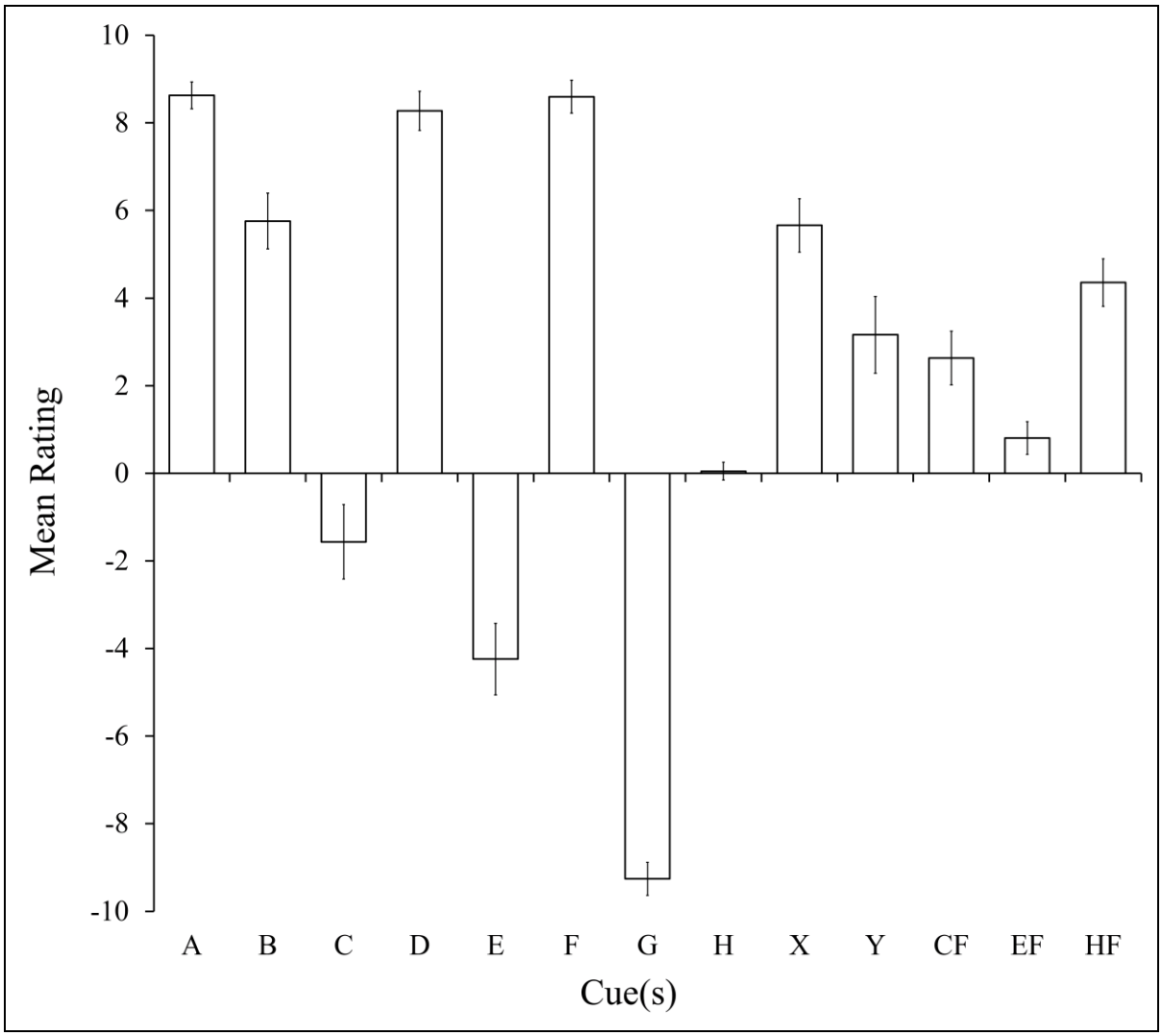

Figure 4. Mean hormone-change ratings at test in Experiment 2 (±SEM). Positive values refer to an increase, zero values to no change, and negative values to a decrease in hormone levels.

To investigate the comparisons between CF, EF, and HF, a one-way ANOVA was conducted on the data. It revealed a significant effect of trial type, $F(2, 60) = 13.08$, $p < .001$, $\eta^2 = .3$. A t-test showed that ratings for EF were lower than for HF, $t(30) = 5.4$, $p < .001$, $d_z =$

.97, indicating that inhibition was obtained in this experiment. In addition, CF had significantly lower ratings than HF, $t(30) = 2.22$, $p = .034$, $d_z = .4$, suggesting that C acquired some inhibitory associative strength. However, EF had lower ratings than CF, $t(30) = 2.82$, $p = .008$, $d_z = .51$, indicating that inhibition for E was greater than for C.

Even though C gained some inhibitory associative strength in this experiment, the redundancy effect was significant. Therefore, it appeared that a lack of inhibition for C was not responsible for the redundancy effect. However, participants varied in the ratings they gave for C: some participants had highly negative ratings, indicating inhibition for this cue, while others rated it at zero, and a small minority of people had positive ratings. It is possible that participants who had negative ratings for C had higher ratings for Y than the participants who did not. To see whether there was a relationship between ratings for C and for Y, a correlation between these two variables was performed. A significant negative correlation between ratings for these cues was found, $r(31) = -.494$, $p = .005$ (Figure 5, left panel). To make sure that higher ratings for Y were not due to general inhibition, we also performed a correlation between ratings for Y and ratings for the inhibitory cue E; this correlation was not significant, $r(31) = .152$, $p = .413$. Given the negative relationship between ratings for C and for Y, it may be expected that participants who had lower ratings for C would also have a smaller redundancy effect. In other words, ratings for C would be positively correlated with the magnitude of the redundancy effect (calculated as X ratings – Y ratings). However, while this correlation was positive, it did not reach significance, $r(31) = .242$, $p = .19$ (Figure 5, right panel).

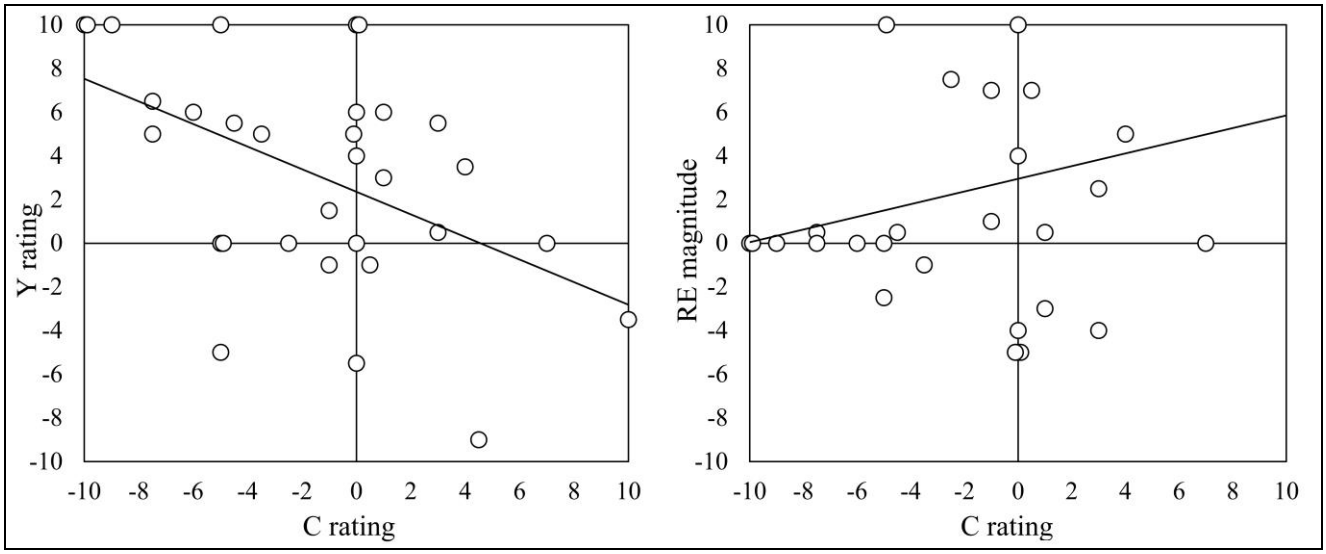

Figure 5. Left panel: Negative correlation between ratings for C and for Y. Right panel: Positive (non-significant) correlation between ratings for C and the magnitude of the redundancy effect (X ratings – Y ratings).

Discussion

Experiment 2 used a task in which the consumption of different medicines led to an increase, no change, or a decrease in hormone levels. This type of task was aimed to more closely match the symmetrical inhibition-excitation continuum assumed by the Rescorla-Wagner (1972) model. In this task inhibitory cues could have independent negative effects on the outcome and caused decreases in hormone levels. Firstly, we investigated whether evidence of inhibition would be obtained in this task and in particular, inhibition for C, which was predicted to become a weak inhibitor. We found that evidence for inhibition was obtained: a cue compound which included an inhibitory cue E (EF) had lower ratings than the cue compound which included a neutral cue H (HF). This highlights that it was possible to obtain inhibition in this task. We also found that C gained some inhibitory associative strength: CF had lower ratings than HF. However, C was not as strong an inhibitor as E: EF had lower ratings than CF. Secondly, we explored whether the redundancy effect in this task

would be reversed, with greater causal ratings for Y than for X, as per Rescorla-Wagner model's predictions. We found that the redundancy effect was significant, even though C was shown to have gained some inhibitory associative strength. This is in contrast to the Rescorla-Wagner model, which predicts greater positive associative strength for Y than for X when C is inhibitory. Notably however, there was individual variability in participants' ratings, and ratings for C and Y correlated negatively: lower ratings for C were related to higher ratings for Y. This provides evidence for one part of the Rescorla-Wagner model's prediction, that in a BY+/CY^0 trial-arrangement, inhibition for C will be related to excitation for Y. While a relationship between ratings for C and the magnitude of the redundancy effect may have been expected, this correlation did not reach significance. It is possible that even though ratings for C and for Y were related, this did not affect the magnitude of the redundancy effect. However, it is also possible that the limited sample size (N = 31) in this experiment was the reason that this correlation did not reach significance. To test the latter possibility, the next experiment explored whether a significant correlation between ratings for C and the magnitude of the redundancy effect would be observed with a greater number of participants.

Experiment 3

In Experiment 3 we used the same task as in the previous experiment to investigate whether the correlation between ratings for C and the magnitude of the redundancy effect would reach significance with a greater number of participants. Given that findings of Experiment 2 verified that inhibition could be obtained in the hormone task, in Experiment 3 we simplified the experimental design, limiting it only to the cues of interest. The design consisted of A+/AX+/BY+/CY^0/D- trials; D- trials were included to ensure that participants saw evidence that single inhibitory cues could have independent effects on the outcome. The design of Experiment 3 is presented in Table 3. At test, we included a cue-compound AD which was a simplified test for the presence of inhibition. The compound AD was included to

check whether participants understood that a cue which led to an increase (A) and a cue which led to a decrease (D) would lead to no change in hormone levels when presented together. We expected ratings for AD to be at zero for participants who learned this.

Table 3.

The design of Experiment 3.

Stage 1	Test
A+	A
AX+	B
BY+	C
CY^0	D
D-	X
	Y
	AD
x 12	x 2

Method

Participants. Participants were 50 University of Plymouth students aged 18-34 years (M = 19.68, SD = 3) and nine were male.

Materials. The materials and procedure in Experiment 3 were the same as in Experiment 2 unless otherwise stated.

The cues were six images of different colour medicines: blue, green, orange, pink, purple, and yellow. These images were randomly assigned to each type of cue (A, B, C, D, X, Y) for each participant.

Procedure. In Stage 1 participants were presented with 12 blocks of trials with the five trial types (A+/AX+/BY+/CY^0/D-) appearing once per block in a random order, with no successive repetitions.

Participants rated all of the individual medicines at test twice, followed by AD trials.

Results

Stage 1. Participants learned the pairings in Stage 1 (Figure 6). In the final epoch they responded correctly on 98.6% (*SD* = 8.27%) of the trials.

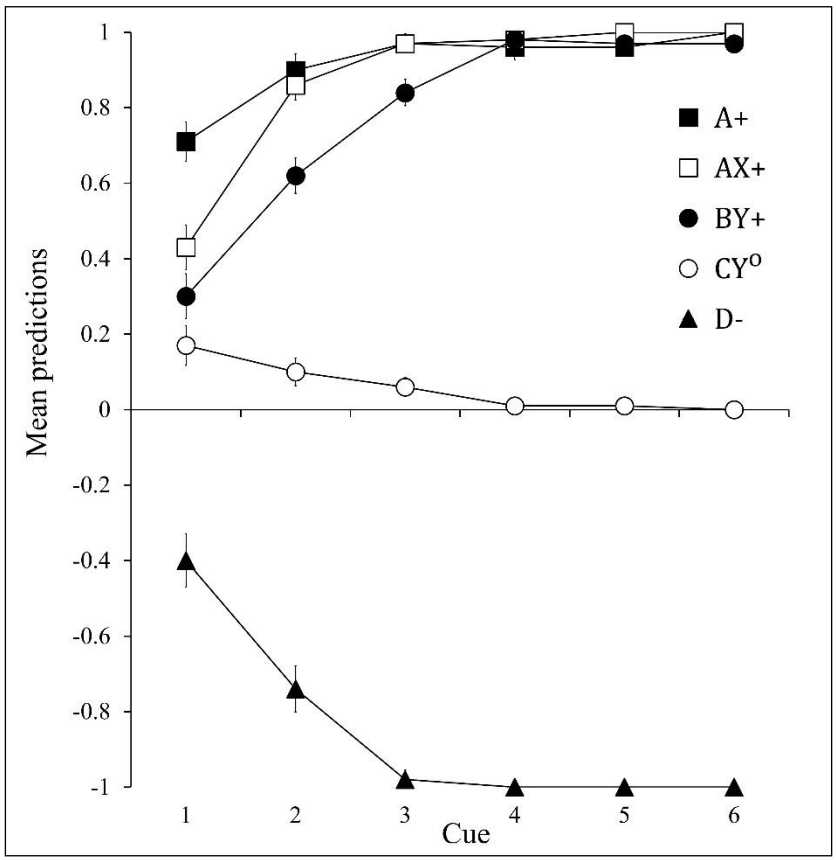

Figure 6. Mean hormone-change predictions throughout the six epochs of Stage 1 in Experiment 3 (±SEM). Positive values refer to a prediction of an increase, zero values refer to a prediction of no change, and negative values refer to a prediction of a decrease in hormone levels.

Test. Figure 7 shows ratings for the cues at test, averaged across participants. A one-way ANOVA for the effect of cue (A, B, C, D, X, Y) revealed a significant main effect, $F(2.42, 118.67) = 185.94$, $p < .001$, $\eta^2 = .79$. Bonferroni-corrected paired comparisons indicated that ratings for B, X, and Y did not differ from each other, ts $\leqslant 2.66$, ps $\geqslant .159$, d_zs $\leqslant .38$, but ratings for all other cues did, ts $\geqslant 7.09$, ps $< .001$, d_zs ≥ 1. A t-test showed that the redundancy effect was not obtained, $t(49) = .96$, $p = .343$, $d_z = .14$, $BF_{01} = 4.22$. Since in this experiment we were interested in exploring the relationships between ratings for C, ratings for Y, and the magnitude of the redundancy effect, the redundancy effect failing to reach significance was surprising, but did not prevent us from exploring these.

Mean ratings for the cue-compound AD were close to zero, averaging at 0.34 (SD = 1.53) and did not significantly differ from zero as shown by a one-sample t-test, $t(49) = 1.57$, $p = .123$, $d_z = .22$, $BF_{01} = 2.08$). All but four participants rated AD as zero and the four participants who did not had positive ratings (1, 2, 4, 10). In Stage 1, A was shown to increase, and D was shown to decrease hormone levels; AD ratings close to zero indicated that participants understood that these cues presented together would lead to no change in hormone levels, providing some evidence of inhibition. Rerunning the analyses without the four participants who had AD ratings as greater than zero did not change the pattern of results detailed in this experiment, therefore these are unreported.

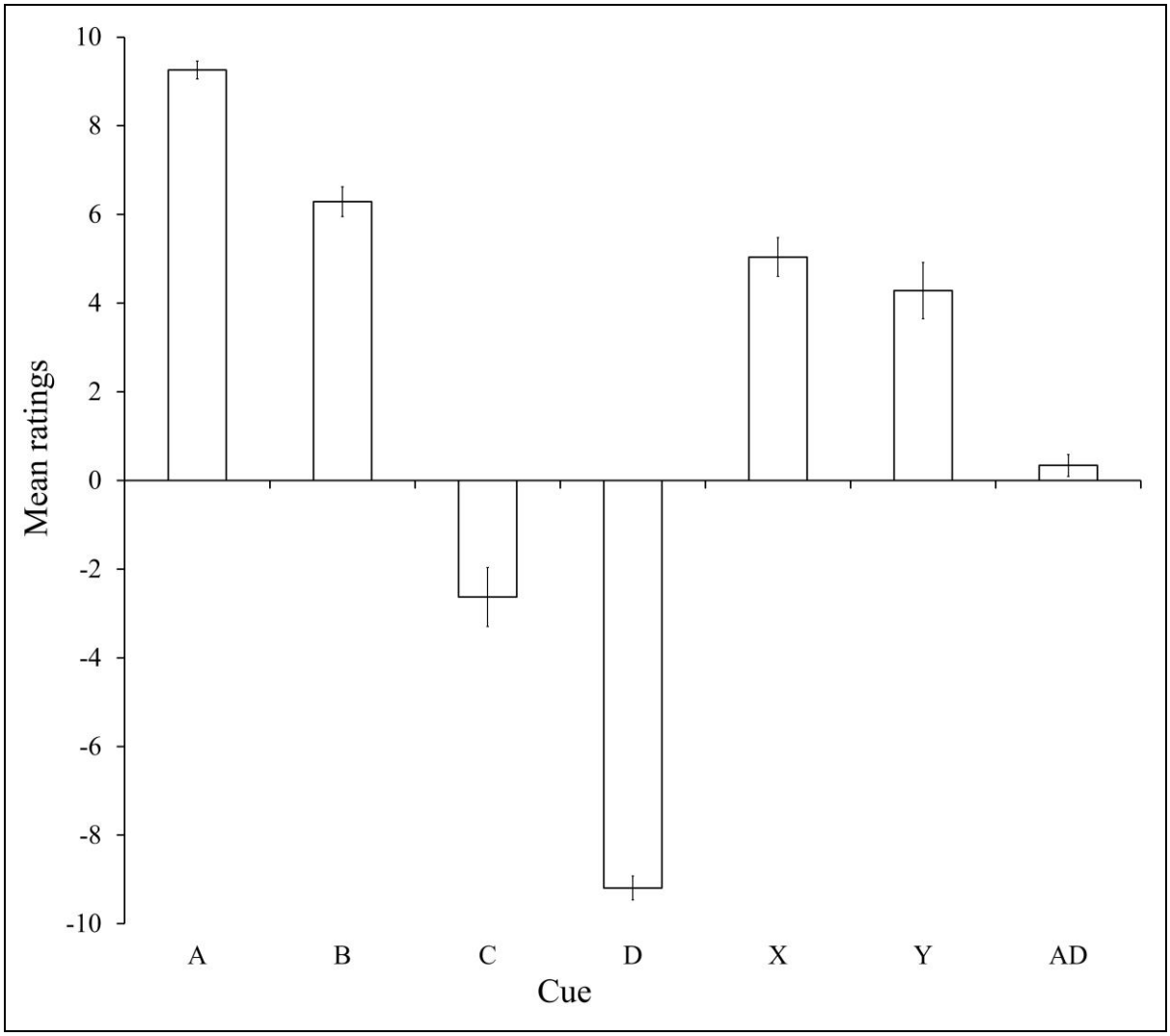

Figure 7. Mean hormone-change ratings at test in Experiment 3 (±SEM). Positive values refer to an increase, zero values to no change, and negative values to a decrease in hormone levels.

Correlations. Once again there was a significant negative correlation between ratings for C and for Y, $r(50) = -.645$, $p < .001$ (Figure 8, left panel). There was also a significant positive correlation between ratings for C and the magnitude of the redundancy effect, $r(50) = .518$, $p < .001$ (Figure 8, right panel).

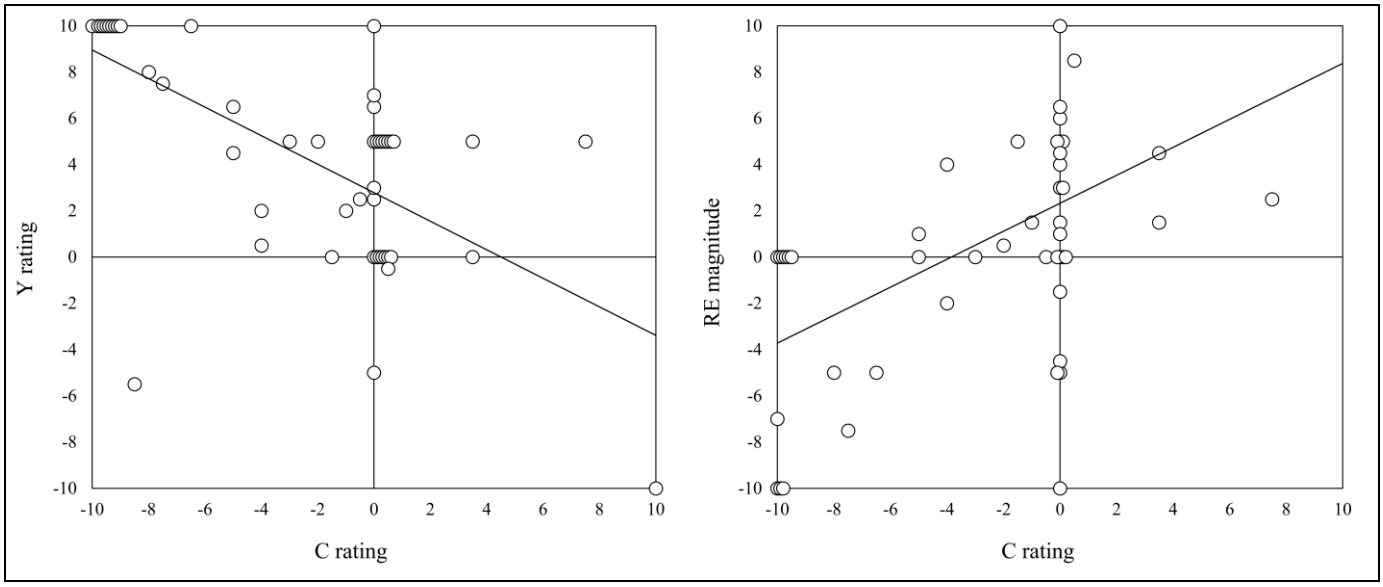

Figure 8. Correlations in Experiment 3. Left panel: Negative correlation between ratings for C and for Y. Right panel: Positive correlation between ratings for C and the magnitude of the redundancy effect (X ratings – Y ratings).

Discussion

In Experiment 3 we set out to explore whether the relationship between ratings for C and the magnitude of the redundancy effect would strengthen with a greater number of participants. Firstly, we replicated the significant negative correlation between ratings for C and for Y. Secondly, a significant positive correlation between ratings for C and the magnitude of the redundancy effect was observed. Overall, findings of this experiment indicated that the redundancy effect was related to the extent to which C was rated as inhibitory: greater inhibition for C was related to a smaller redundancy effect.

Interestingly, the redundancy effect was not observed in this experiment. It is not immediately clear why it was not significant in this experiment but significant in Experiment 2; both used the same task. However, in this experiment mean ratings for C were lower (M = -2.63) than in Experiment 2 (M = -1.56). Therefore, individual variation resulting in lower

ratings for C could have been related to higher ratings for Y, and a smaller redundancy effect which did not reach significance in this experiment.

Given the positive relationship between inhibition for C and the magnitude of the redundancy effect, in Experiment 4 we aimed to see whether it was possible to experimentally manipulate participants' causal assumptions about C by establishing it as either inhibitory or neutral, and observe the corresponding differences in the magnitude of the redundancy effect.

Experiment 4

In Experiment 4 we set out to investigate whether ratings for Y and subsequently the redundancy effect could be manipulated by overtly changing the causal nature of C. We hypothesised that demonstrating that C had inhibitory effects on the outcome would result in higher ratings for Y and a smaller redundancy effect than demonstrating that C was neutral and had no effects on the outcome. Both of these interpretations should be possible because they are consistent with the contingencies; participants may assume that on BY+/CY^0 trials both B and Y have excitatory effects on the outcome and lead to an increase in hormone levels while C has inhibitory effects, leading to a decrease in hormone levels. Alternatively, they may assume that B has excitatory effects on the outcome, leading to an increase in hormone levels, while C and Y have neutral effects and lead to no change. Experiment 4 proceeded as follows. In Stage 1 participants were presented with A+/AX+/BY+/CY^0/D- trials as in the previous experiment. Following the completion of Stage-1 training, they were asked to provide ratings for each individual cue (Test 1). In Stage 2, participants were shown the same trial types as in Stage 1, but with additional trials on which C was presented alone. For participants in the inhibitory group, C was shown to lead to a decrease (C-), while for participants in the neutral group, C was shown to lead to no change (C^0) in hormone levels.

Subsequently to this manipulation, participants were asked to provide ratings for each cue once again (Test 2). The full design of Experiment 4 is shown in Table 4. We expected that the redundancy effect would be smaller in the inhibitory group than in the neutral group at Test 2.

Table 4.

The design of Experiment 4. In Stage 2 participants in the inhibitory group were presented with C- trials while participants in the neutral group were presented with C^0 trials.

Stage 1	Test 1	Stage 2	Test 2
A+	A	A+	A
AX+	B	AX+	B
BY+	C	BY+	C
CY^0	D	$C-/C^0$	D
D-	X	CY^0	X
	Y	D-	Y
x 12	x 2	x 8	x 2

Method

Participants. Participants were 70 University of Plymouth students aged 18-41 (M = 20.64, SD = 4.09) and 11 were male. There were 34 participants in the inhibitory group and 36 participants in the neutral group.

Materials. The materials and procedure in Experiment 4 were the same as in the previous experiment unless otherwise stated.

Images of the medicines were: blue, green, orange, pink, purple, and yellow. The medicines were randomly assigned to each type of cue (A, B, C, D, X, Y) for each participant.

Procedure. In Stage 1 participants were presented with 12 blocks of trials with the five trial types (A+/AX+/BY+/CY^0/D-) appearing once per block. Subsequently, participants rated all of the individual medicines twice at Test 1. In Stage 2 participants were presented with eight blocks of trials, including trials on which C was presented alone. The inhibitory group were presented with A+/AX+/BY+/C-/CY^0/D-, while the neutral group were presented with A+/AX+/BY+/C^0/CY^0/D-. At Test 2 participants were asked to rate each cue again, twice. The trial types within each block were presented in a random order, with no successive repetitions of the same trial type between blocks.

Results

For the key analyses regarding the redundancy effect, please see subsection "The redundancy effect based on group".

Stage 1. Participants learned the contingencies in Stage 1. They responded correctly in the final epoch on 99.57% (*SD* = 4.62%) of the trials (inhibitory group: *M* = 99.71%, *SD* = 3.83%; neutral group: *M* = 99.44%, *SD* = 5.26%). These data are shown in the upper panels of Figure 9. A two-way ANOVA using the variables of trial type (A+, AX+, BY+, CY^0, D-) and group (inhibitory vs neutral) on the final epoch of Stage-1 predictions indicated a significant effect of trial type, $F(4, 272) = 26046.09$, $p < .001$, $\eta_P^2 > .99$, no significant effect of group, $F(1, 68) = .35$, $p = .558$, $\eta_P^2 < .01$, and no significant interaction, $F(4, 272) = .31$, $p = .872$, $\eta_P^2 < .01$.

Stage 2. Lower panels of Figure 9 show predicted changes in hormone levels in Stage 2 for participants in the inhibitory group (left panel) and the neutral group (right panel). In the

final epoch of Stage 2 correct responses were made on 99.93% (*SD* = 8.74%) of the trials (inhibitory group: *M* = 98.04%, *SD* = 12%; neutral group: *M* = 99.77%, *SD* = 3.4%). A two-way ANOVA using the variables of trial type (A+, AX+, BY+, C-/C^0, CY^0, D-) and group (inhibitory vs neutral) on the final epoch of Stage-2 predictions revealed a significant effect of trial type, $F(5, 340) = 6454.23$, $p < .001$, $\eta_P^2 = .99$, a significant effect of group, $F(1, 68) = 643.75$, $p < .001$, $\eta_P^2 = .9$, and a significant interaction, $F(5, 340) = 330.09$, $p < .001$, $\eta_P^2 = .83$. Importantly, the only significant differences between the groups were observed on C-alone trials. Consistently with the manipulation, a greater number of participants in the inhibitory group predicted a decrease in hormone levels than the neutral group, $t(33) = 67$, $p < .001$, $d_s = 16.02$. Predictions for the other cues did not differ significantly between the groups, ts ≤ 1, ps $\geq .307$, d_ss $\leq .24$.

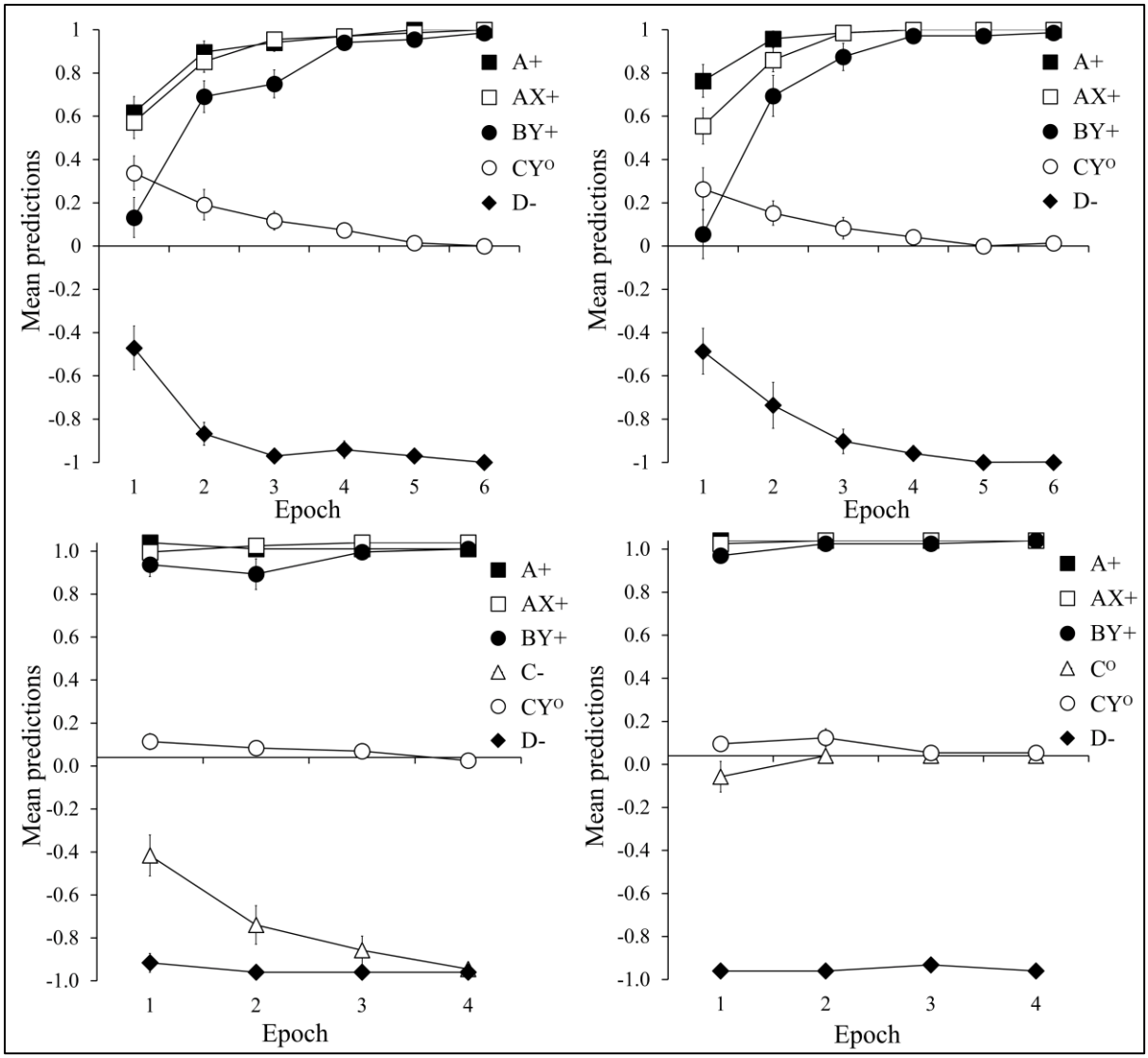

Figure 9. Mean hormone-change predictions in Stage 1 and in Stage 2 in Experiment 4 (± between-subjects SEM). Upper left panel: Stage-1 responses in the inhibitory group. Upper right panel: Stage-1 responses in the neutral group. Lower left panel: Stage-2 responses in the inhibitory group. Lower right panel: Stage-2 responses in the neutral group. Positive values refer to a prediction of an increase, zero values refer to a prediction of no change, and negative values refer to a prediction of a decrease in hormone levels.

Test. Figure 10 shows the mean hormone-change ratings in both groups at Test 1 (left panel) and at Test 2 (right panel).

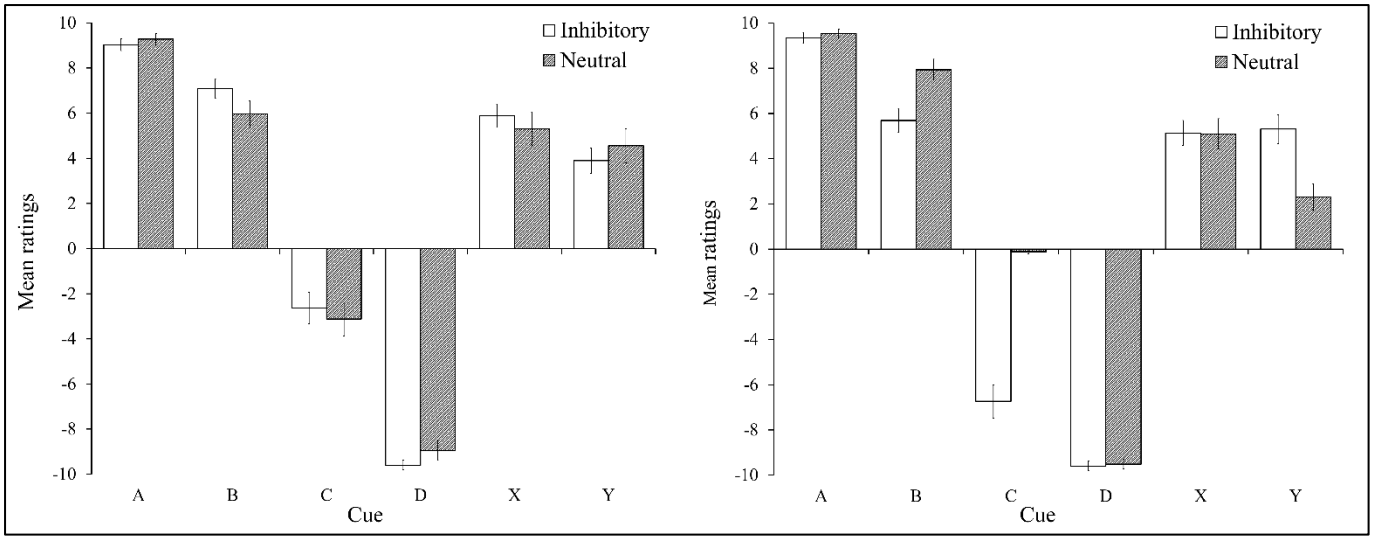

Figure 10. Mean hormone-change ratings in Experiment 4 (± between-subjects SEM). Left panel: Responses from Test 1. Right panel: Responses from Test 2. Positive values refer to an increase, zero values to no change, and negative values to a decrease in hormone levels.

In order to make sure that there were no differences in ratings between the groups at Test 1, a two-way ANOVA for the variables of cue (A, B, C, D, X, Y) and group (inhibitory vs neutral) was conducted on the data. It revealed a significant effect of cue, $F(3.04, 206.66) = 314.93$, $p < .001$, $\eta_P^2 = .82$. Bonferroni-corrected paired comparisons indicated that ratings between B and X, and between X and Y, did not differ significantly, ts $\leqslant 2.57$, ps $\geqslant .167$, d_zs $\leqslant .31$, while ratings between all other cues did, ts $\geqslant 3.98$, ps $\leqslant .002$, d_zs $\geqslant .48$. There was no significant effect of group, $F(1, 68) = .14$, $p = .714$, $\eta_P^2 < .01$, and no significant interaction, $F(5, 340) = .88$, $p = .496$, $\eta_P^2 = .02$.

In order to check whether establishing C as either inhibitory or neutral affected ratings for the cues at Test 2, a two-way ANOVA using the variables of cue (A, B, C, D, X, Y) and group (inhibitory vs neutral) was conducted. It revealed a significant effect of cue, $F(3.83, 260.47) = 464.98$, $p < .001$, $\eta_P^2 = .87$, a significant effect of group, $F(1, 68) = 13.78$, $p < .001$,

η_P^2 = .17, and a significant interaction, $F(3.83, 260.47) = 23.63$, $p < .001$, η_P^2 = .26. Simple main effects analyses indicated that the groups did not differ in their ratings for A, D, and X, ts ≤ .68, ps ≥ .5, d_ss ≤ .16, but the inhibitory group had lower ratings for B, $t(68) = 3.18$, p = .002, d_s = .76, lower ratings for C, $t(33.83) = 8.95$, $p < .001$, d_s = 2.14, and higher ratings for Y, $t(68) = 3.57$, p = .001, d_s = .85, than the neutral group. Group differences in ratings for C and for Y were consistent with our predictions. While differences for B were not predicted, it is possible to explain how inhibition for C resulted in lower ratings for B in the inhibitory group. Establishing C as inhibitory led to higher ratings for Y. Because B and Y were presented together on BY+ trials, it is likely that the stronger relationship between Y and the outcome would have decreased the strength of the causal relationship between B and the outcome in this group.

The redundancy effect based on group. A three-way ANOVA using the variables of cue (X vs Y), group (inhibitory vs neutral), and test (1 vs 2) revealed a significant three-way interaction, $F(1, 68) = 14.09$, $p < .001$, η_P^2 = .17. To explore this interaction, 2 x 2-way ANOVA tests for cue (X vs Y) and group (inhibitory vs neutral) were conducted on the ratings at Test 1 and Test 2.

The two-way ANOVA on Test-1 data revealed a significant effect of cue, $F(1, 68) = 6.81$, p = .011, η_P^2 = .09, with higher ratings for X than for Y. There was no significant effect of group, $F(1, 68) = .002$, p = .965, η_P^2 < .001, and no significant interaction, $F(1, 68) = 1.44$, p = .235, η_P^2 = .02.

The two-way ANOVA on Test-2 data revealed a significant effect of cue, $F(1, 68) = 5.8$, p = .019, η_P^2 = .08; ratings for X were higher than for Y. There was also a significant effect of group, $F(1, 68) = 5.38$, p = .023, η_P^2 = .07, and a significant interaction, $F(1, 68) = 7.47$, p = .008, η_P^2 = .1. Simple main effects analyses indicated that the redundancy effect was

observed in the neutral group, $t(35) = 3.8$, $p = .001$, $d_z = .63$, but not in the inhibitory group, $t(33) = .22$, $p = .827$, $d_z = .04$, $BF_{01} = 5.32$.

Discussion

In Experiment 4 we set out to explore whether experimentally establishing the causal status of C as either neutral or inhibitory would affect the magnitude of the redundancy effect. We found that this manipulation had an effect on ratings for Y; they were higher when C was established as inhibitory than when C was established as neutral. As a result, establishing C as inhibitory resulted in a smaller redundancy effect than establishing C as neutral, confirming our predictions. It was also interesting to note that ratings for B were lower in the inhibitory group as a result of the manipulation. Establishing C as an inhibitor, resulted in higher ratings for Y, and therefore it is not surprising that ratings for B were affected, as B was presented with Y on BY+ trials. One interpretation is that in this group, because on BY+ trials Y was more causal, the strength of the excitatory relationship between B and the outcome would be reduced. If this interpretation is correct, it would suggest that participants appeared to treat the effects of cues on the outcome as additive, showing cue competition. One other factor which may have influenced the reduced ratings for B is that C was established as a strong inhibitor. Rescorla-Wagner (1972) model predicts that C will become a weak inhibitor while in Experiment 4 we established it as a strong inhibitory cue. It is possible that a manipulation resulting in a weaker inhibition for C would increase the ratings for Y, decrease the ratings for B, and reduce the redundancy effect to a lesser extent than it did in our experiment.

General discussion

In this manuscript we set out to explore whether the redundancy effect could be due to a lack of inhibition for cue C in the design A+/AX+/BY+/CY^0. Rescorla-Wagner (1972) model predicts a stronger association with the outcome for Y than for X in this design, contrary to the observed results. Crucially, this prediction relies on C gaining some inhibitory associative strength, which is predicted to protect Y from extinction. However, there are reasons to doubt that participants have learned that C was inhibitory in the previous demonstrations of the redundancy effect. In these tasks the outcome varied only unidirectionally; it could be either present or absent. Melchers et al. (2006; see also Baetu & Baker, 2010; Lotz & Lachnit, 2009) argued that tasks in which reinforcers cannot take on negative values, do not accurately reflect the assumptions of the Rescorla-Wagner model regarding inhibition. Therefore, if a task was chosen which was a better match for the assumptions of the model, results more in line with its predictions may be observed. In Experiment 1 we employed a task with only neutral and positive outcomes used in the previous research, to explore whether the redundancy effect and inhibition for C could be demonstrated in the same experiment. We observed the redundancy effect but found no evidence of inhibition for C. Therefore, the possibility remained that a lack of inhibition for C may have contributed to the redundancy effect in this task. In Experiment 2, we explored whether inhibition for C and the redundancy effect would be observed in an alternative scenario, which we hypothesised should better reflect the Rescorla-Wagner model's assumptions regarding inhibition. The outcome in this task was the level of a fictional hormone which could increase, not change, or decrease, representing excitatory, neutral, and inhibitory effects on the outcome, respectively. We encouraged learning that single cues could have inhibitory effects on the outcome both by instruction and by overtly presenting cues which led to a decrease in the outcome. We found that inhibition for C and the

redundancy effect were observed in this experiment, in an apparent contradiction to the predictions of the Rescorla-Wagner model. However, there were also individual differences in participants' ratings. Most notably, we found a negative correlation between ratings for C and for Y, indicating that the extent to which C was rated as inhibitory was related to the extent that Y was rated as excitatory. In Experiment 3, with a greater sample of participants, we further found a positive correlation between ratings for C and the magnitude of the redundancy effect, indicating that lower ratings for C were related to a smaller redundancy effect. In Experiment 4, we directly manipulated the causal status of C to be either inhibitory or neutral, and found the corresponding changes in the redundancy effect, consistently with predictions derived from Experiment 3. The redundancy effect was smaller in the group in which C was established as inhibitory than in the group in which C was established as neutral.

Taken together, results from these experiments indicated that the magnitude of the redundancy effect was related to the extent that C was rated as inhibitory. Inhibition for C increased ratings for Y and reduced the magnitude of the redundancy effect. Conversely, neutral causal status of C was related to lower ratings for Y and a larger redundancy effect. In Experiment 4, no redundancy effect was observed in the inhibitory group, indicating that a lack of inhibition is sufficient for the redundancy effect to occur and as such, it is possible that the previous demonstrations of the redundancy effect in this manuscript and elsewhere using the two-outcome task (Jones et al., 2019; Jones & Zaksaite, 2018; Uengoer et al., 2013; Uengoer et al., 2017; Zaksaite & Jones, 2017) were at least partly attributable to a lack of inhibition for C. If C did not become inhibitory, then Y was not protected from extinction, resulting in lower ratings for Y and contributing to the redundancy effect. Data in this manuscript were also more consistent with the predictions of the Rescorla-Wagner (1972) model, as the magnitude of the redundancy effect was reduced with greater inhibition for C.

This pattern of results could be explained by the hormone task better reflecting the symmetry between excitation and inhibition assumed by the model, in line with arguments by Melchers et al. (2006). Therefore, other researchers may wish to consider using such a task when testing predictions of the Rescorla-Wagner model which rely on cues accruing inhibitory associative strength, particularly when inhibition is predicted to be weak.

The finding that ratings for Y were dependent on the causal status of C is particularly interesting in light of earlier demonstrations of protection from extinction in human causal learning (e.g. Holmes, Griffiths, & Westbrook, 2014). While Holmes et al. demonstrated protection from extinction, they found that test ratings of the target cue were not determined by the causal status of its partner, during the preceding extinction trials. Our results are therefore more consistent with the account of protection from extinction offered by the Rescorla-Wagner (1972) model. It is notable that the scenario used by Holmes et al. was similar to that used in Experiment 1 here, where we did not find any evidence that C protected Y from extinction by becoming inhibitory.

These findings may also be relevant for another well-known effect within associative learning: the relative validity effect (Wasserman, 1990; for analogous results in rats see Wagner, Logan, Haberlandt, & Price, 1968 and Wasserman, 1974 in pigeons). In Wasserman's (1990) study, an allergist task was used in which participants were asked to judge the predictiveness of two unique cues presented in compound with one common cue. There were two key conditions in this experiment. In both conditions, the common cue was reinforced 50% of the time. In the first condition, none of the cues predicted the outcome reliably (BY±/CY±). In the second condition, the unique cues predicted the outcome better than the common cue (BY+/CY-). Wasserman observed that the common cue Y was judged to be less predictive of the outcome in the second condition, where other, more reliable predictors of the outcome were present. The contingencies used in the second condition are

identical to the treatment of cue Y in the redundancy effect design. Therefore, it is possible that the redundancy effect and the relative validity effect may rely on common mechanisms. To our knowledge, the relative validity effect in humans has only been demonstrated using a binary outcome task, similar to the allergist task. Given that we found that inhibition for C reduced ratings for cue Y, it is possible that using a task in which inhibitors could have observable effects as in the present studies, would affect the magnitude of the relative validity effect as well.

While these results bring us closer to reconciling the redundancy effect with the predictions of the Rescorla-Wagner (1972) model, additional assumptions are needed for the model to account for our results. The model predicts that if C does not gain inhibitory associative strength, Y will not be protected from extinction, in which case both X and Y should have associative strength close to zero, once learning has reached asymptote. In Experiment 4, in which C was shown to have neutral effects on the outcome, the redundancy effect was still observed, in an apparent contradiction to these predictions. When C is inhibitory, this model predicts a stronger relationship between Y and the outcome than X and the outcome. Establishing C as inhibitory in Experiment 4 did not enable us to obtain the predicted result either, with equivalent causal ratings observed for X and for Y. One way to reconcile the model with our data is to assume that learning about X had not reached asymptote, with the consequence that X retained some associative strength acquired as a result of its pairing with the outcome on AX+ trials. Another possibility, however, is that X was rated as a moderately likely cause of the outcome because its effects on the outcome are uncertain. Jones, Zaksaite, and Mitchell (2019) used confidence ratings and base rate manipulations to show that uncertainty about the effects that X has on the outcome contributes to the redundancy effect. A detailed exploration of this idea is beyond the scope of this article, but one explanation of the redundancy effect is that it is due not only to low

ratings for Y (in the absence of protection from extinction by C), but also intermediate ratings for X as a result of uncertainty about its causal status.

While models which use a global error-term, like the Rescorla-Wagner (1972) model, struggle to predict the redundancy effect, single error-term models can predict the redundancy effect because X is followed by the outcome more often than Y. However, they do not predict cue competition effects such as blocking without incorporating an extra process. One model which incorporates both is Mackintosh's (1975) theory of selective attention. It contains a single error-term enabling the prediction that X will have a stronger association with the outcome than Y. It also predicts that cues presented on the same trial will compete for increases in associability, which enables this theory to explain cue competition effects. While associability for both X and Y is predicted to decrease as they are presented with other cues which are more informative about the occurrence of the outcome (A, B, and C), it is difficult to predict for which cue this decline will be faster. One possibility is that the decrease in associability is faster for X because its companion, A, was shown to be the perfect predictor of the outcome. An alternative possibility is that the decline in associability for Y is faster, because it was presented twice as often as X, as well as having been paired with other cues that are perfect predictors of both the presence (B), and the absence (C), of the outcome. Due to the lack of specification of this theory it is not possible to accurately predict how attention interacts with previous associative history to determine learning, but, provided that any decline in associability for X does not exceed that for Y, this theory can account for the redundancy effect.

However, if Mackintosh's theory could explain the redundancy effect, then we might expect differences in attention between X and Y. This was investigated using eye-tracking by Jones and Zaksaite (2018) who did not observe any differences in visual attention between these cues. Furthermore, Uengoer et al. (2017) did not find evidence for differences in

associability between X and Y, although it is possible that any differences may have been too small to detect using these procedures. It is also unfortunate, however, that Mackintosh's theory of selective attention has difficulty accounting for the results of this manuscript as it does not easily explain the development of conditioned inhibition (although see Moore & Stickney, 1985; Schmajuk & Moore, 1985).

Another model which includes a single error-term and an additional process to explain cue competition effects is the comparator hypothesis (Denniston, Savastano, & Miller, 2001; Miller & Matzel, 1988; Stout & Miller, 2007). This theory proposes that the strength of the association between a cue and the outcome is based on single error-term rules. At the point of performance, when participants are asked to estimate the extent to which a cue causes the outcome, a comparator process takes place. This involves a comparison between the associative strength for the target cue (e.g. X) with the associative strength for any companion cues the target was presented with during learning (e.g. A). If the companion cue has a strong association with the outcome, this reduces the strength of the response for the target cue. On the other hand, if the companion cue has a weak association with the outcome then the strength of the response for the target cue is increased. Applied to the redundancy effect, this theory predicts that even though X accrues positive associative strength during learning because of single error-term rules, at test responding for X will be low, because it was presented with A, and A had a strong association with the outcome. Cue Y on the other hand, was presented with two other cues, one of which had a strong association with the outcome (B). However, Y starts with lower associative strength because it was paired with the outcome only 50% of the time. Therefore, higher responding for X than for Y is predicted. While this theory can account for the redundancy effect, Experiment 1 of Jones and Pearce (2015), Experiment 2 of Uengoer et al., (2013), and Zaksaite and Jones (2017) obtained results that challenge this theory.

The findings of this manuscript, taken together with the earlier findings by Jones et al. (2019), suggest that the redundancy effect may be multiply-determined: a lack of inhibition for C contributes to low ratings for Y, and participants' uncertainty about the causal status of X contributes to high ratings for X. Future studies are invited to explore whether the redundancy effect can be reversed by using both manipulations in one experiment, one which encourages inhibition for C and another which aims to resolve participants' uncertainty about X. It is also important to note that there may be other factors still to be discovered, which relate to the redundancy effect. For example, while Experiment 1 of Uengoer et al. (2013) and Experiment 1 of Jones et al. (2019) showed that the redundancy effect was not due to a failure of blocking, Experiments 2, 3, and 4 of this manuscript used a different scenario and had no control cue for blocking. Therefore, it would be interesting for further studies to explore whether blocking is obtained in this scenario.

One useful approach to further illuminate the reasons behind the redundancy effect may be considering the role of individual differences. In Experiments 2, 3, and 4, participants varied in their ratings for C and Y, and consequently the magnitude of the redundancy effect. One reason for this may be that participants had different assumptions about the effects that C and Y had on the outcome. For example, the assumption that C as well as Y led to no change in the outcome, and the assumption that C led to a decrease and Y led to an increase, would both have been consistent with the contingencies of the task. In the former case B could have been seen as causing an increase, with both C and Y having neutral effects. In the latter case, B and Y could have been assumed to lead to an increase, while C led to a decrease. It is also interesting to note that assuming that B led to an increase while C and Y had neutral effects would be more consistent with the predictions of single error-term models, while assuming that B and Y led to an increase and C led to a decrease would be more in line with predictions of summed error-term models. Following this line of reasoning, there may be individual

differences in the extent to which participants rely on rules consistent with single and summed error-term models. Hybrid models of learning, which include both single and summed error-terms, do exist (e.g. Le Pelley, 2004), however these assume that properties of the cues determine the extent to which each will be utilised, with little scope for the inclusion of systematic individual differences. Even though it has been argued that an approach incorporating individual differences in learning may be needed to understand the full complexity of learning and behaviour (e.g. Byrom, 2013; Sauce & Matzel, 2013), to our knowledge, variation in the extent to which participants utilise processes consistent with single and summed error has not been considered in published work to date.

Conclusion

In this manuscript we investigated whether a lack of inhibition for cue C contributed to the redundancy effect. We used a task in which inhibitors were shown to have independent effects on the outcome, which we hypothesised would better reflect the assumptions of the Rescorla-Wagner (1972) model regarding inhibition than the allergist task. We found a negative relationship between ratings for C and for Y: greater inhibition for C was related to higher ratings for Y and a smaller redundancy effect. In Experiment 4, this link was evidenced via an experimental manipulation; when C was established as an inhibitor, the redundancy effect was smaller than when C was established as neutral. Therefore, the previous demonstrations of the redundancy effect which used the allergist task (Jones et al., 2019; Jones & Zaksaite, 2018; Uengoer et al., 2013; Zaksaite & Jones, 2017) may have been due to a lack of inhibition for C. Future studies are suggested to explore the extent to which the redundancy effect is multiply-determined, and individual differences regarding the redundancy effect, particularly differences in the extent to which people systematically utilise rules consistent with single or summed error. Researchers may also wish to use a task which

is more consistent with the assumptions of the Rescorla-Wagner model when testing its predictions which relate to inhibition, particularly weak inhibition.

References

Aitken, M. R. F., Larkin, M. J., & Dickinson, A. (2000). Super-learning of causal judgements. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *53*(1), 59–81.

Baetu, I., & Baker, A. G. (2010). Extinction and blocking of conditioned inhibition in human causal learning. *Learning & Behavior, 38*(4), 394-407.

Beckers, T., De Houwer, J., Pineño, O., & Miller, R. R. (2005). Outcome additivity and outcome maximality influence cue competition in human causal learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 238–249.

Byrom, N. C. (2013). Accounting for individual differences in human associative learning. *Frontiers in Psychology*, *4*, 1–8.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42-45.

Denniston, J. C., Savastano, H. I., & Miller, R. R. (2001). The extended comparator hypothesis: Learning by contiguity, responding by relative strength. In R. R. Mowrer & S. B. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 65–117). London, UK: Lawrence Erlbaum Associates.

Detke, M. J. (1991). Extinction of sequential conditioned inhibition. *Animal Learning & Behavior*, *19*, 345-354.

Dickinson, A., Shanks, D., & Evenden, J. (1984). Judgement of act-outcome contingency: The role of selective attribution. *Quarterly Journal of Experimental Psychology, 36A*(January), 29–50.

Holland, P. C. (1985). The nature of conditioned inhibition in serial and simultaneous feature negative discriminations. In R. R. Miller & N. E. Spear (Eds.), *Information processing in animals: Conditioned inhibition* (pp. 267-297). Hillsdale, NJ: Erlbaum.

Holmes, N. M., Griffiths, O., & Westbrook, R. F. (2014). The influence of partner cues on the extinction of causal judgements in people. *Learning & Behavior, 42*(3), 289-303.

JASP Team (2018). JASP (Version 0.9)[Computer software].

Jeffreys, H. (1961). *The theory of probability* (3rd ed.). Oxford: Oxford University Press.

Jones, P. M., & Pearce, J. M. (2015). The fate of redundant cues: Further analysis of the redundancy effect. *Learning & Behavior, 43*, 72–82.

Jones, P. M., & Zaksaite, T. (2018). The redundancy effect in human causal learning: No evidence for changes in selective attention. *The Quarterly Journal of Experimental Psychology, 71*(8), 1748-1760.

Jones, P. M., Zaksaite, T., & Mitchell, C. J. (2019). Uncertainty and blocking in human causal learning. *Journal of Experimental Psychology: Animal Learning and Cognition, 45*(1), 111-124.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behaviour* (pp. 279–296). New York: Appleton Century Crofts.

Konorski, J. (1948). Conditioned reflexes and neuron organization. New York, NY, US: Cambridge University Press.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer fort-tests and ANOVAs. *Frontiers in Psychology, 4*, 1-12.

Larkin, M. J., Aitken, M. R., & Dickinson, A. (1998). Retrospective revaluation of causal judgments under positive and negative contingencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1331-1352.

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *57*, 193–243.

Livesey, E. J., & Boakes, R. A. (2004). Outcome additivity, elemental processing and blocking in human causality judgements. *The Quarterly Journal of Experimental Psychology*, *57B*(4), 361–379.

Lotz, A., & Lachnit, H. (2009). Extinction of conditioned inhibition: Effects of different outcome continua. *Learning & Behavior*, *37*(1), 85-94.

Lovibond, P. E., Been, S. L., Mitchell, C. J., Bouton, M. E., & Frohardt, R. (2003). Forward and backward blocking of causal judgment is enhanced by additivity of effect magnitude. *Memory & Cognition*, *31*(1), 133–142.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*(4), 276–298.

Melchers, K., Lachnit, H., & Shanks, D. (2004). Within-compound associations in retrospective revaluation and in direct learning: A challenge for comparator theory. *The Quarterly Journal of Experimental Psychology: Section B, 57*(1), 25-53.

Melchers, K. G., Wolff, S., & Lachnit, H. (2006). Extinction of conditioned inhibition through nonreinforced presentation of the inhibitor. *Psychonomic Bulletin & Review*, *13*(4), 662–7.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla–Wagner model. *Psychological Bulletin*, *117*, 363-386.

Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 51–92). San Diego, CA: Academic Press.

Miller, R. R., & Schachtman, T. R. (1985). Conditioning context as an associative baseline: Implications for response generation and the nature of conditioned inhibition. In R. R. Miller & N. E. Spear (Eds.), *Information processing in animals: Conditioned inhibition* (pp. 51-88). Hillsdale, NJ: Erlbaum.

Mitchell, C. J., & Lovibond, P. F. (2002). Backward and forward blocking in human electrodermal conditioning: Blocking requires an assumption of outcome additivity. *The Quarterly Journal of Experimental Psychology: Section B, 55*(4), 311–329.

Mitchell, C. J., Lovibond, P. F., & Condoleon, M. (2005). Evidence for deductive reasoning in blocking of causal judgments. *Learning and Motivation, 36*(1), 77–87.

Moore, J. W., & Stickney, K. J. (1985). Antiassociations: Conditioned inhibition in attentional–associative networks. In R. R. Miller & N. E. Spear (Eds.), *Information*

processing in animals: Conditioned inhibition. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex*. Oxford: Oxford University Press.

Pearce, J. M., Nicholas, D. J., & Dickinson, A. (1982). Loss of associability by a conditioned inhibitor. *Quarterly Journal of Experimental Psychology*, *34*(3), 149-162.

Pearce, J. M., Dopson, J. C., Haselgrove, M., & Esber, G. O. R. (2012). The fate of redundant cues during blocking and a simple discrimination. *Journal of Experimental Psychology: Animal Behavior Processes*, *38*(2), 167–179.

Rescorla, R. A. (1969). Pavlovian Condition Inhibition. *Psychological Bulletin*, *72*, 77–94.

Rescorla, R. A. (1982). Some consequences of associations between the excitor and the inhibitor in a conditioned inhibition paradigm. *Journal of Experimental Psychology: Animal Behavior Processes*, *8*(3), 288-298.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Projasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, US: Appleton Century Crofts.

Sauce, B., & Matzel, L. D. (2013). The causes of variation in learning and behavior: Why individual differences matter. *Frontiers in Psychology*, *4*(395), 7–16.

Schmajuk, N. A., & Moore, J. W. (1985). Real-time attentional models for classical conditioning and the hippocampus. *Physiological Psychology, 13*, 278–290.

Stout, S. C., & Miller, R. R. (2007). Sometimes-competing retrieval (SOCR): A formalization of the comparator hypothesis. *Psychological Review*, *114*(3), 759–783.

Uengoer, M., Dwyer, D. M., Koenig, S., & Pearce, J. M. (2019). A test for a difference in the associability of blocked and uninformative cues in human predictive learning. *The Quarterly Journal of Experimental Psychology*, *72*(2), 222-237.

Uengoer, M., Lotz, A., & Pearce, J. M. (2013). The fate of redundant cues in human predictive learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *39*(4), 323–33.

Vogel, E. H., & Wagner, A. R. (2017). A theoretical note in interpretation of the "Redundancy Effect" in associative learning. *Journal of Experimental Psychology: Animal Learning and Cognition*, *43*(1), 119–125.

Wagner, A. R., Logan, F. A., & Haberlandt, K. (1968). Stimulus selection in animal discrimination learning. Journal of Experimental Psychology, 76(2p1), 171-180.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, *121*(2), 222–236.

Wasserman, E. A. (1974). Stimulus-reinforcer predictiveness and selective discrimination learning in pigeons. *Journal of Experimental Psychology*, *103*(2), 284-297.

Wasserman, E. A. (1990). Attribution of causality to common and distinctive elements of compound stimuli. *Psychological Science*, *1*(5), 298-302.

Williams, D. A., & Overmier, J. B. (1988). Some types of conditioned inhibitors carry collateral excitatory associations. *Learning & Motivation*, *19*(4), 345-368.

Yarlas, A. S., Cheng, P. W., & Holyoak, K. J. (1995). Alternative approaches to causal induction: The probabilistic contrast versus the Rescorla-Wagner model. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the Seventeenth Annual Meeting of the Cognitive Science Society* (pp. 431–436). NJ: Erlbaum.

Zaksaite, T., & Jones, P. M. (2017). The redundancy effect in human causal learning: Evidence against a Comparator Theory explanation. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*.

Zimmer-Hart, C. L., & Rescorla, R. A. (1974). Extinction of Pavlovian conditioned inhibition. *Journal of Comparative and Physiological Psychology*, *86,* 837–845.